810002

EcoRIa   24   39   54
GAATTCCGG TGC AGG ACG AAG CTG TTC TGG ATT TCT TAC AGT GAT GGG GAC CAG
          Arg Thr Lys Leu Phe Trp Ile Ser Tyr Ser Asp Gly Asp Gln

                    69   84   99
TGT GCC TCA AGT CCA TGC CAG AAT GGG GGC TCC TGC AAG GAC CAG CTC CAG TCC
Cys Ala Ser Ser Pro Cys Gln Asn Gly Gly Ser Cys Lys Asp Gln Leu Gln Ser

      114   129   144   159
TAT ATC TGC TTC TGC CTC CCT GCC TTC GAG GGC CGG AAC TGT GAG ACG CAC AAG
Tyr Ile Cys Phe Cys Leu Pro Ala Phe Glu Gly Arg Asn Cys Glu Thr His Lys

              174   189   204   Pst Ia
GAT GAC CAG CTG ATC TGT GTG AAC GAG AAC GGC GGC TGT GAG CAG TAC TGC AGT
Asp Asp Gln Leu Ile Cys Val Asn Glu Asn Gly Gly Cys Glu Gln Tyr Cys Ser

219   234   249   264
GAC CAC ACG GGC ACC AAG CGC TCC TGT CGG TGC CAC GAG GGG TAC TCT CTG CTG
Asp His Thr Gly Thr Lys Arg Ser Cys Arg Cys His Glu Gly Tyr Ser Leu Leu

          279   294   ACA GTT GAA TAT CCA TGT GGA AAA ATA CCT
GCA GAC GGG GTG TCC TGC ACA CCC ACA GTT GAA TAT CCA TGT GGA AAA ATA CCT
Ala Asp Gly Val Ser Cys Thr Pro Thr Val Glu Tyr Pro Cys Gly Lys Ile Pro

      Xba I   339   354   369
ATT CTA GAA AAA AGA AAT GCC AGC AAA CCC CAA GGC CGA ATT GTG GGG GGC AAG
Ile Leu Glu Lys Arg Asn Ala Ser Lys Pro Gln Gly Arg Ile Val Gly Gly Lys

      384   399   414   429
GTG TGC CCC AAA GGG GAG TGT CCA TGG CAG GTC CTG TTG TTG GTG AAT GGA GCT
Val Cys Pro Lys Gly Glu Cys Pro Trp Gln Val Leu Leu Leu Val Asn Gly Ala

              444   459   474
CAG TTG TGT GGG GGG ACC CTG ATC AAC ACC ATC TGG GTG GTC TCC GCG GCC CAC
Gln Leu Cys Gly Gly Thr Leu Ile Asn Thr Ile Trp Val Val Ser Ala Ala His

489   504   519   534
TGT TTC GAC AAA ATC AAG AAC TGG AGG AAC CTG ATC GCG GTG CTG GGC GAG CAC
Cys Phe Asp Lys Ile Lys Asn Trp Arg Asn Leu Ile Ala Val Leu Gly Glu His

          549   564   579   594
GAC CTC AGC GAG CAC GAC GGG GAT GAG CAG AGC CGG CGG GTG GCG CAG GTC ATC
Asp Leu Ser Glu His Asp Gly Asp Glu Gln Ser Arg Arg Val Ala Gln Val Ile

                  609   Sma I   624   639
ATC CCC AGC ACG TAC GTC CCG GGC ACC ACC AAC CAC GAC ATC GCG CTG CTC CGC
Ile Pro Ser Thr Tyr Val Pro Gly Thr Thr Asn His Asp Ile Ala Leu Leu Arg

      654   669   684   699
CTG CAC CAG CCC GTG GTC CTC ACT GAC CAT GTG GTG CCC CTC TGC CTG CCC GAA
Leu His Gln Pro Val Val Leu Thr Asp His Val Val Pro Leu Cys Leu Pro Glu

FIG. 1A

```
                          -60                      -50                        -60↓
                MetValSerGlnAlaLeuArgLeuLeuCysLeuLeuLeuGlyLeuGlyGlyLeuAlaLeuAlaGlyGlyValAlaLeuAlaSerGly
TCAACAGCCACGGGCACCCAGCCACTGCACAGACATTTCATCATGGTCTCCCAGGCCCTTCAGGCTGCTCTGCTTCTTCTTCTTGGCTTCAGGGCTGGCTAGGGCCCGGGCTGGCTAAGGCCTCAGGA
      10        20        30        40        50        60        70        80        90       100       110       120
         -30                   -20                       -10                    -1 +1
   GlyGluThrArgAspMetProTrpLysProGlyProMetLysArgValPheValThrGlnLeuGlnLeuAsnIleGlyValLeuHisAlaArgArgArgAlaAsnAlaPheLeuGluLeuArg
   GGACAAACACGGGACATGCCGTGGAAGCCGGGCCTCACGAGTCTTTGTAAACCAGGAGAAGCCCACGGCGTCCTGCACGGGGCGCGGCGGCCAGCGGTTCCTGGAGGAGCTGCG
      130       140       150       160       170       180       190       200       210       220       230       240
      +10                  +20                       +30                        +40
   ProGlySerLeuGluArgGluCysLysGluGluLnCysSerPheGluGluAlaArgGluIlePheLysAspAlaGluArgThrLysLeuPheTrpIleSerTyrSerAspGlyAspGln
   CCCGGGCTCCCTGCACAGGGACTGCAAGGAGGAGCAGTGCTCCTTCGAGGAGGCCCGGGAGATCTTCAAGGACGCCGAGAGGACAAAGCTGTTCTGGATTTCTTACAGTGATGGGGACCAG
      250       260       270       280       290       300       310       320       330       340       350       360
      +50                  +60                       +70                        +80
   CysAlaSerSerProCysGlnAsnGlyGlySerCysLysAspGlnLeuGlnSerTyrIleCysPheCysLeuProAlaPheGluGlyArgAsnCysGluThrHisLysAspAspGlnLeu
   TGTGCCTCAAGTCCATGCCAGAATGGGGGCTCCTGCAAGGACCAGCTGCAGTCCTATATCTGCTTCTGCCTGGCCTTCGAGGGCCGGAACTGTGAGACGCACAAGGATGACCAGCTG
      370       380       390       400       410       420       430       440       450       460       470       480
      +90                  +100                      +110                       +120
   IleCysValSerGlnAlaAsnGlyGlyCysGluGlnTyrCysSerAspHisThrGlyThrLysArgSerCysArgCysHisGluGlyTyrSerLeuLeuAlaAspGlyValSerCysThrPro
   ATCTGTGTCAGCCAAGCCGGCTGTGAGCAGTACTGCAGTGACCACAGGGCACCAAGCGCTCCTGTGGCTGCCACGAGGGTTACTCTCTGCTGGCAGACGGGGTGTCCTGCACACCC
      490       500       510       520       530       540       550       560       570       580       590       600
      +130                 +140                      +150                       +160
   ThrValGluTyrProCysGlyLysIleProIleLeuGluLysArgAsnAlaSerLysProGlnGlyArgIleValGlyGlyLysValCysProLysGlyGluCysProTrpGlnValLeu
   ACAGTTGAATATCCATGTGGAAAAATACCTATTCTAGAAAAAAGAAATGCCAGCAAACCCCAAGGCCGAATTGTCGGGGCAAGGTGTGCCCCAAAGGGGAGTGTCCATGGCAGGTCCTC
      610       620       630       640       650       660       670       680       690       700       710       720
      +170                 +180                      +190                       +200
   LeuLeuValAsnGlyAlaGlnLeuCysGlyGlyThrLeuIleAsnThrIleTrpValValSerAlaAlaHisCysPheAspLysIleLysAsnTrpArgAsnLeuIleAlaValLeuGly
   TTGTTGCTGAATGGAGCTCAGTTGTTGGGGGGACCCTGATCAACACCATCTGGGTGGTCTCGGCGGCCCACTGTTTCGACAAAATCAAGAACTGGAGGAACCTGATCGCCGTGCTGGGC
      730       740       750       760       770       780       790       800       810       820       830       840
      +210                 +220                      +230                       +240
   GluHisAspLeuSerGluHisAspAspGlnAspLeuSerSerArgArgValAlaGlnIleIleIleProSerThrTyrValProGlyThrThrAsnHisAspIleAlaLeuLeuArgLeuHis
   GAGCACGACCTCAGCGAGCACGACGGGCAGTGAGCAGAGCGGCGGGTGCGCCAGGTCATCATCCCCAGCACGTACGTGCCGGGCACCACCAACCACGACATCGCGCTGCTCCGGCTGCAC
      850       860       870       880       890       900       910       920       930       940       950       960
      +250                 +260                      +270                       +280
   GlnProValValLeuThrAspHisValValProLeuCysLeuProGluArgThrPheSerGluArgThrLeuAlaPheValArgPheSerLeuValSerGlyTrpGlyGlnLeuLeuAsp
   CAGCCCGTGGTCCTCACCGACCATGTGGTGCCCCTCTGCCTGCCGGAGCGGACGTTCTCTGAGAGGACGCTGGCCTTCGTGCGGTTCTCACTGGTCTCAGGCTGGGGCCAGCTGCTGGAC
      970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
      +290                 +300                      +310                       +320
   ArgGlyAlaThrAlaLeuGluLeuMetValLeuAsnValProArgLeuMetThrGlnAspCysLeuGlnGlnSerArgLysValGlyAspSerProAsnIleThrGluTyrMetPheCys
   CGTGGCGCCACGGCCCTGGAGCTCATGGTCCTCAACGTCCCCCGGCTCATGACCCAGGACTGCCTGCAGCAGTCACGGAAGGTGGGAGACTCCCCAAATATCACGGAGTACATGTTCTGT
     1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
      +330                 +340                      +350                       +360
   AlaGlyTyrSerAspGlySerLysAspSerCysLysGlyAspSerGlyGlyProHisAlaThrHisTyrArgGlyThrTrpTyrLeuThrGlyIleValSerTrpGlyGlnGlyCysAla
   GCCGGCTACTGGATGGCAGCAAGGACTCCTGCAAGGGGGACAGCTGGAGGCCCCACATGCCACCCACTACCGGGGCACGTGGTACCTGACGGGCATCGTCAGCTGGGGCCAGGGCTGCGCA
     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
      +370                 +380                      +390           +400        +406
   ThrValGlyHisPheGlyValTyrThrArgValSerGlnTyrIleGluTrpLeuGlnLysLeuMetArgSerGluProArgProGlyValLeuLeuArgAlaProPhePro***
   ACCGTTGGCCACTTTGGGGTGTACACCAGGGTCTCCCAGTACATGGAGTGGCTGCAAAAGCTCATGCGCTCAGAGCCCAGGCCAGGCGTCCTGCTCCGAGCCCCATTTCCCTAGCCCAGC
     1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440

AGCCCTGGCCTGTGGAGAGAAAGCCAAGGCTGCGTCGAACTGTCCTGGCACCAAATCCCATATATTTCTTCTGCACTTAATGGGGTAGAGGAGGGCCATGGGAGGGAGGGCAGAGGTGGGCAG
     1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

GGCACACACACAGAAACACAGACACACAGAGACACAGAGACACTCAGGGAGACACTCTCAGCACATCGCACAGACTCAAAGCAGACTCCAAGCATTCAAAGCAGCTAATACAGACACACAGA
     1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680

GATGGAATACAAAACATCAGACGCCAGACGCCAGCACGGCCTGCACGAGAGGGGCCAGGGGAGTGCCAAGGTTGTCCTGGAGGCAGACAGCCCAGCTCAGCCTCCTTACCTCCCTTCAGCCAA
     1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800

GCCCCACCTGCACGTGATCTGCTGGCCCTCAGGCTGCTGCTCTGCCTTCATTGCTGGAGCAGTAGAGGCATGAACACACATGGATGCACACACACACGCCAATGCACACACACAGAG
     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920

ATATGCACACACACGGATGCACACACAGATGGTCACACAGAGATACGCAAACACACGGATGCACACGCACATAGAGATATGCACACACAGTGCACACACAGATATACACATGGATGCAC
     1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040

GCACATGCCCAATGCACGCCACACATCAGTGCACACGGATGCACACAGAATATGCACACACGGATGTGCGCACACACAGATATGCACACACAGATATGCACACACACATGGATGAGCACACACACCAAGTGCGCACA
     2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160

CACACCGGATGTACACACACAGATGCACACACAGATGCACACACACCGGATGCTGACTCCATGTGTGCTCTCCTCTGAAGGCGGTTGTTTAGCTCTCACTTTTCTGGTTCTTATCCATTATC
     2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

ATCTTCACTTCAGACAATTCAGAAGGCATCACCATGCATGGTGGCGAATGCCCCCAAACTCTCCCCCAAATGTATTTCTCCCTTGCTGTGGTGCCGGCCTGCACAGACTATTCCCCACCTG
     2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

CTTCCCCAGCTTCACAATAAAACGGCTGCGTCTCCTCCGCCACACCTGTGGTGGCTGCACCCAAAAAAAAAAAAAAAAAAAAAA
     2410      2420      2430      2440      2450      2460      2470      2480
```

FIG.1B

```
                    10        20        30        40
          1          *       * *        *     *        *    *
FACTOR VII    ANA-FLYYLRPGSLYRYCKYYQCSFYYARYIFYXKXRTKL
FACTOR IX     YNSGKLYYFVVQGNLYRYCMYYKCSFYYARYVPYFYNTYRTTY
FACTOR X      ANS-FLYYMKKGHLYRYCMYYTCSYYYARYVPYDSDKTNY
PROTEIN C     ANS-FLYYLRHSSLYRYCIYYICDFYYALYIFQNVDDTLA
PROTHROMBIN   ANT-FLYYVRKGNLYRYCVYYTCSYYYAFYALYSSTATDV

                  50              60              70
FACTOR VII    FWISYSDGDQCASS------PCQNGGSCKDQLQSYICF
FACTOR IX     FWKQYVDGDQCESN------PCLNGGSCKBDINSYECW
FACTOR X      FWNKYKDGDQCETS------PCQNQGKCKBGCLGEYTCT
PROTEIN C     FWSKHVDGDQCLVLPLEHPCASLCCCHGTCIBGICSFSCD
PROTHROMBIN   FWARYTACETARTPRDKLAACLEGNCAEGLGTNYRGHVNI

                80              90             100            110
FACTOR VII    CLPAFEGRNCETHKDDQLICVNENGGCEEQYCSDHTGTKRSC
FACTOR IX     CPPFGFEGKNCELDVT----CNIKNGRCEQFCKNSADNKVVC
FACTOR X      CLEGFEGKNCELFTRKL--CSLDNGDCDQFCHEEQNS-VVC
PROTEIN C     CRSGWEGRFCQREVSFLN-CSLDNGGCTHYCLEEVGW-RRC
PROTHROMBIN   TRSGIECQLWRSRYPHKP-EINSTTHPGADLQENFCRNPDS

                120             130            140
FACTOR VII    RCHEGYSLLADGVSCTPTVEYPCGKIPILEKRNASKPQQGR
FACTOR IX     SCTEGYRLAGNQKSCEPAVPFPCGRVSVSQTSKLRT
FACTOR X      SCARGYTLADNGKACIPTGPYPCGKQTLER
PROTEIN C     SCAPGYKLGDDLLQCHPAVKFPCGRPWKRMEKKRSHL
PROTHROMBIN   SNTGPWCYTTDPTVRRQECSIPVCGQDQVTVAMTPRS
```

FIG. 2A

810002

```
                    1        10        20        30    36
From cDNA          X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X X R T
                           * *       * *       * *  * *  *   *    *
Amino Acid Sequence        A N A F L Y Y L R P G S L Y R Y C K Y Y Q C S F Y Y A R Y I F Y X X X X X

                    40        50        60        70
cDNA               K L F W I S Y S D G D Q C A S S P C Q N G G S C K D Q L Q S Y I C P C L
Amino Acid Sequence  L F W I S Y S D G D Q C A S S P C Q N G G S C K D Q L Q     I C P C L

                    80        90        100
cDNA               P A F E G R N C E T H K D D Q L I C V N E N G G C E Q Y C S D H T G T K
Amino Acid Sequence P A F E G R N C E T H K D D Q L                       C S D H T G T

                    110       120       130       140
cDNA               R S C R C H E G Y S L L A D G V S C T P T V E Y P C G K I P I L E K R N
Amino Acid Sequence  S C R C H E G Y S L L A D G V S C T P T V E Y                 E K R ( )

                    150
cDNA               A S K P Q G R
Amino Acid Sequence A S K P Q G R
```

FIG. 2B

FIG. 4

810002
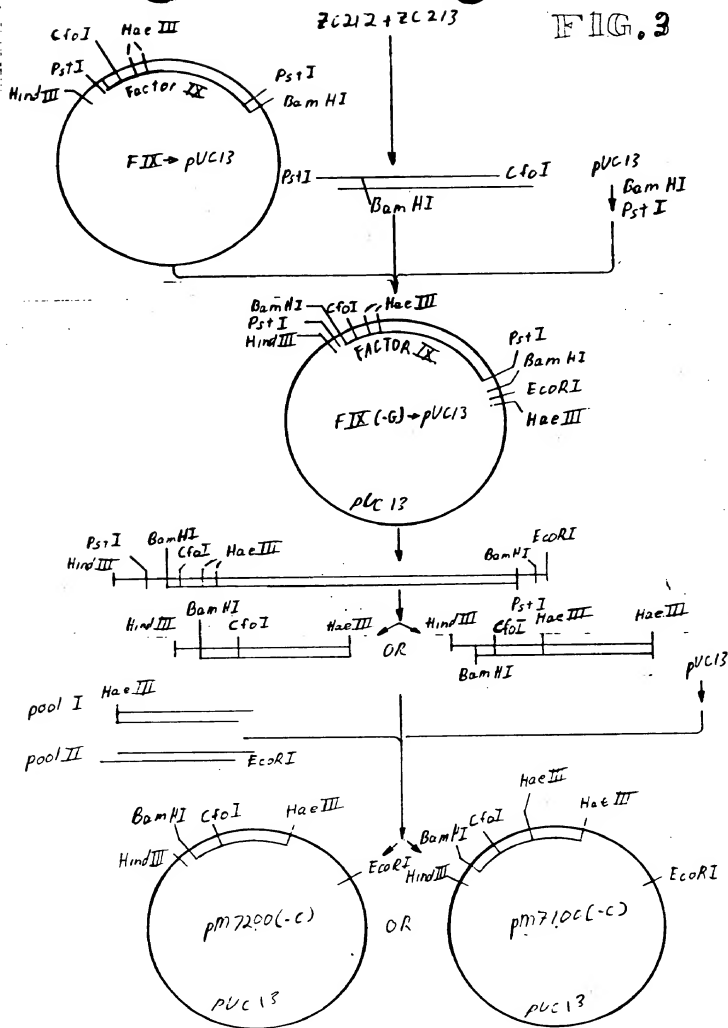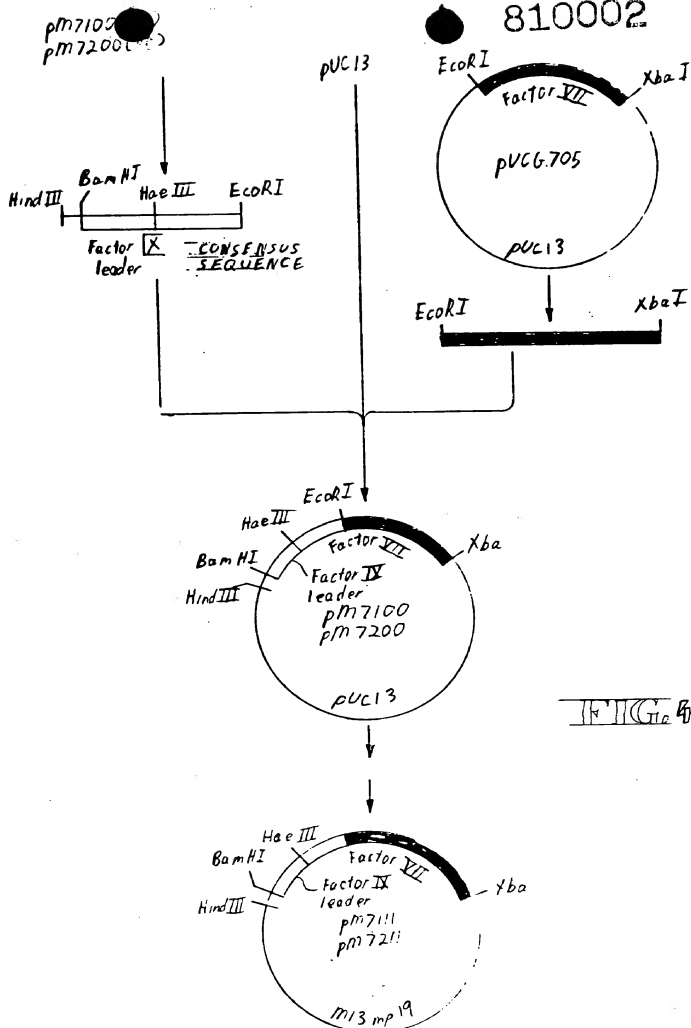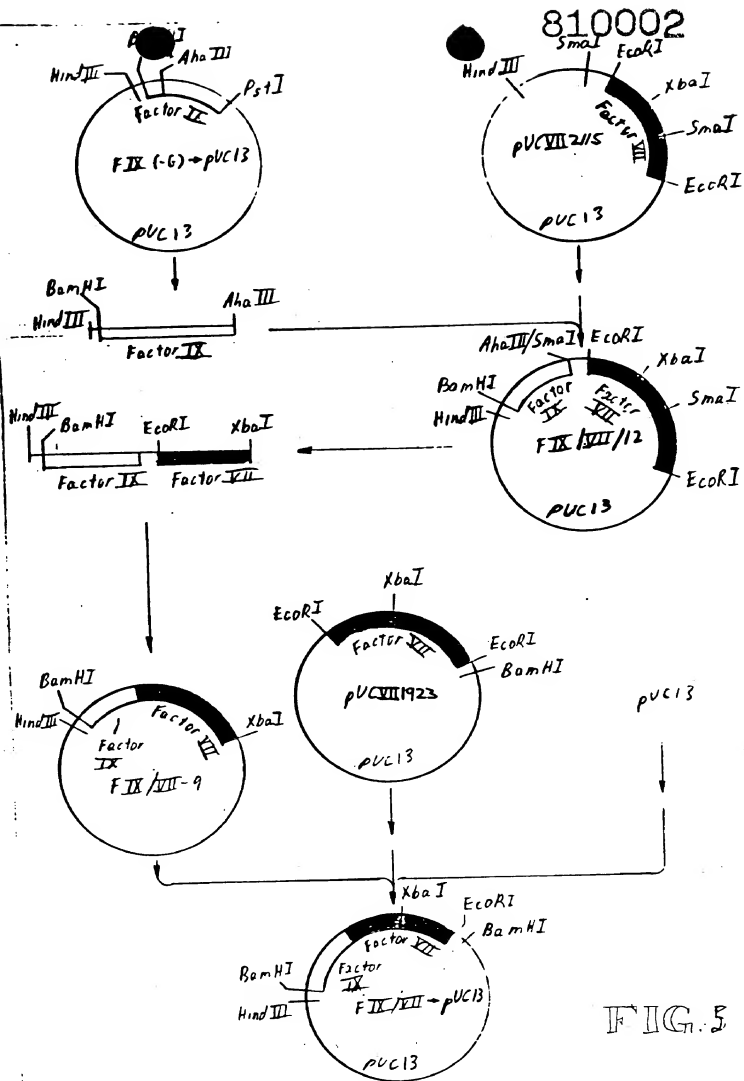
FIG.5
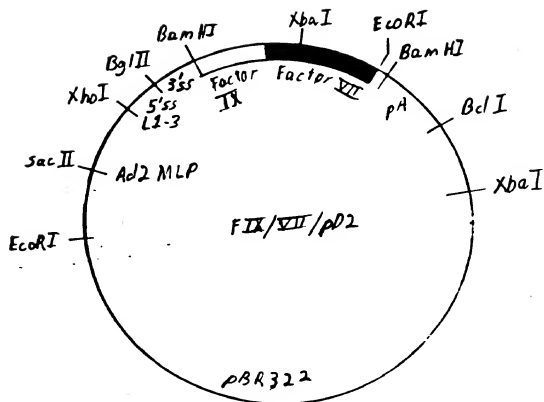
FIG. 6
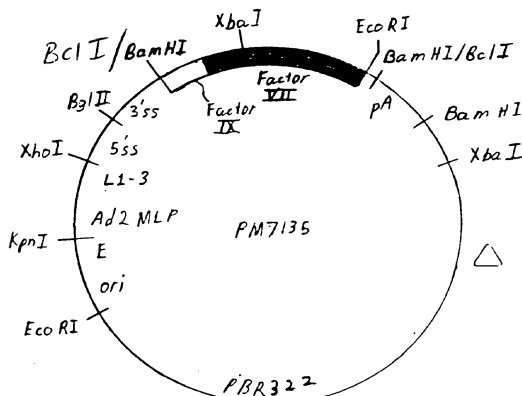
FIG. 8

FIGURE 7

```
                                21                        36
GGATCC ATG CAG CGC GTG AAC ATG ATC ATG GCA GAA TCA CCA GGC
       MET Gln Arg Val Asn MET Ile MET Ala Glu Ser Pro Gly


                                66                        81
CTC ATC ACC ATC TGC CTT TTA GGA TAT CTA CTC AGT GCT GAA TGT
Leu Ile Thr Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys


    96                         111                       126
ACA GTT TTT CTT GAT CAT GAA AAC GCC AAC AAA ATT CTG AAT CGG
Thr Val Phe Leu Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg


    141                        156                       171
CCA AAG AGG TAT AAT TCA GGT AAA TTG GAA GAG TTT GTT CAA GGG
Pro Lys Arg Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly


    186                        201                       216
AAC CTT GAG AGA GAA TGT ATG GAA GAA AAG TGT AGT TTT GAA GAA
Asn Leu Glu Arg Glu Cys MET Glu Glu Lys Cys Ser Phe Glu Glu


    231                        246                       261
GCA CGA GAA GTT TTT GAA AAC ACT GAA AGA ACA AAG CTG TTC TGG
Ala Arg Glu Val Phe Glu Asn Thr Glu Arg Thr Lys Leu Phe Trp


    276                        291                       306
ATT TCT TAC AGT GAT GGG GAC CAG TGT GCC TCA AGT CCA TGC CAG
Ile Ser Tyr Ser Asp Gly Asp Gln Cys Ala Ser Ser Pro Cys Gln


    321                        336                       351
AAT GGG GGC TCC TGC AAG GAC CAG CTC CAG TCC TAT ATC TGC TTC
Asn Gly Gly Ser Cys Lys Asp Gln Leu Gln Ser Tyr Ile Cys Phe


    366                        381                       396
TGC CTC CCT GCC TTC GAG GGC CGG AAC TGT GAG ACG CAC AAG GAT
Cys Leu Pro Ala Phe Glu Gly Arg Asn Cys Glu Thr His Lys Asp


    411                        426                       441
GAC CAG CTG ATC TGT GTG AAC GAG AAC GGC GGC TGT GAG CAG TAC
Asp Gln Leu Ile Cys Val Asn Glu Asn Gly Gly Cys Glu Gln Tyr


    456                        471                       486
TGC AGT GAC CAC ACG GGC ACC AAG CGC TCC TGT CGG TGC CAC GAG
Cys Ser Asp His Thr Gly Thr Lys Arg Ser Cys Arg Cys His Glu


    501                        516                       531
GGG TAC TCT CTG CTG GCA GAC GGG GTG TCC TGC ACA CCC ACA GTT
Gly Tyr Ser Leu Leu Ala Asp Gly Val Ser Cys Thr Pro Thr Val


    546                        561                       576
GAA TAT CCA TCT GGA AAA ATA CCT ATT CTA GAA AAA AGA AAT GCC
Glu Tyr Pro Cys Gly Lys Ile Pro Ile Leu Glu Lys Arg Asn Ala


    591                        606                       621
AGC AAA CCC CAA GGC CGA ATT GTG GGG GGC AAG GTG TGC CCC AAA
Ser Lys Pro Gln Gly Arg Ile Val Gly Gly Lys Val Cys Pro Lys
```

```
     636                      651                      666
GGG GAG TGT CCA TGG CAG GTC CTG TTG TTG GTG AAT GGA GCT CAG
Gly Glu Cys Pro Trp Gln Val Leu Leu Leu Val Asn Gly Ala Gln

     681                      696                      711
TTG TGT GGG GGG ACC CTG ATC AAC ACC ATC TGG GTG GTC TCC GCG
Leu Cys Gly Gly Thr Leu Ile Asn Thr Ile Trp Val Val Ser Ala

     726                      741                      756
GCC CAC TGT TTC GAC AAA ATC AAG AAC TGG AGG AAC CTG ATC GCG
Ala His Cys Phe Asp Lys Ile Lys Asn Trp Arg Asn Leu Ile Ala

     771                      786                      801
GTG CTG GGC GAG CAC GAC CTC AGC GAG CAC GAC GGG GAT GAG CAG
Val Leu Gly Glu His Asp Leu Ser Glu His Asp Gly Asp Glu Gln

     816                      831                      846
AGC CGG CGG GTG GCG CAG GTC ATC ATC CCC AGC ACG TAC GTC CCG
Ser Arg Arg Val Ala Gln Val Ile Ile Pro Ser Thr Tyr Val Pro

     861                      876                      891
GGC ACC ACC AAC CAC GAC ATC GCG CTG CTC CGC CTG CAC CAG CCC
Gly Thr Thr Asn His Asp Ile Ala Leu Leu Arg Leu His Gln Pro

     906                      921                      936
GTG GTC CTC ACT GAC CAT GTG GTG CCC CTC TGC CTG CCC GAA CGG
Val Val Leu Thr Asp His Val Val Pro Leu Cys Leu Pro Glu Arg

     951                      966                      981
ACG TTC TCT GAG AGG ACG CTG GCC TTC GTG CGC TTC TCA TTG GTC
Thr Phe Ser Glu Arg Thr Leu Ala Phe Val Arg Phe Ser Leu Val

     996                     1011                     1026
AGC GGC TGG GGC CAG CTG CTG GAC CGT GGC GCC ACG GCC CTG GAG
Ser Gly Trp Gly Gln Leu Leu Asp Arg Gly Ala Thr Ala Leu Glu

    1041                     1056          ,          1071
CTC ATG GTC CTC AAC GTG CCC CGG CTG ATG ACC CAG GAC TGC CTG
Leu MET Val Leu Asn Val Pro Arg Leu MET Thr Gln Asp Cys Leu

    1086                     1101                     1116
CAG CAG TCA CGG AAG GTG GGA GAC TCC CCA AAT ATC ACG GAG TAC
Gln Gln Ser Arg Lys Val Gly Asp Ser Pro Asn Ile Thr Glu Tyr

    1131                     1146                     1161
ATG TTC TGT GCC GGC TAC TCG GAT GGC AGC AAG GAC TCC TGC AAG
MET Phe Cys Ala Gly Tyr Ser Asp Gly Ser Lys Asp Ser Cys Lys

    1176                     1191                     1206
GGG GAC AGT GGA GGC CCA CAT GCC ACC CAC TAC CGG GGC ACG TGG
Gly Asp Ser Gly Gly Pro His Ala Thr His Tyr Arg Gly Thr Trp

    1221                     1236                     1251
TAC CTG ACG GGC ATC GTC AGC TGG GGC CAG GGC TGC GCA ACC GTG
Tyr Leu Thr gly Ile Val Ser Trp Gly Gln Gly Cys Ala Thr Val
```

```
    1266                   1281                      1296
GGC CAC TTT GGG GTG TAC ACC AGG GTC TCC CAG TAC ATC GAG TGG
Gly His Phe Gly Val Tyr Thr Arg Val Ser Gln Tyr Ile Glu Trp

    1311                   1326                      1341
CTG CAA AAG CTC ATG TCA GAG CCA CGC CCA GGA GTC CTC CTG
Leu Gln Lys Leu MET Arg Ser Glu Pro Arg Pro Gly Val Leu Leu

    1356                        1378      1388       1398
CGA GCC CCA TTT CCC TAG CCCAGCAGCC CTGGCCTGTG GAGAGAAAGC
Arg Ala Pro Phe Pro   .

         1408        1418       1428        1438        1448
CAAGGCTGCG TCGAACTGTC CTGGCACCAA ATCCCATATA TTCTTCTGCA


         1458        1468       1478        1488        1498
GTTAATGGGG TAGAGGAGGG CATGGGAGGG AGGGAGAGGT GGGGAGGGAG


         1508        1518       1528        1538        1548
ACAGAGACAG AAACAGAGAG AGACAGAGAC AGAGAGAGAC TGAGGGAGAG


         1558        1568       1578        1588        1598
ACTCTGAGGA CCATGGAGAG AGACTCAAAG AGACTCCAAG ATTCAAAGAG


         1608        1618       1628        1638        1648
ACTAATAGAG ACACAGAGAT GGAATAGAAA AGATGAGAGG CAGAGGCAGA


         1658        1668       1678        1688        1698
CAGGCGCTGG ACAGAGGGGC AGGGGAGTGC CAAGGTTGTC CTGGAGGCAG


         1708        1718       1728        1738        1748
ACAGCCCAGC TGAGCCTCCT TACCTCCCTT CAGCCAAGCC CCACCTGCAC


         1758        1768       1778        1788        1798
GTGATCTGCT GGCCCTCAGG CTGCTGCTCT GCCTTCATTG CTGGAGACAG


         1808        1818       1828        1838        1848
TAGAGGCATG ACACACATGG ATGCACACAC ACACACGCCA TGCACACACA


         1858        1868       1878        1888        1898
CAGAGATATG CACACACACG GATGCACACA CAGATGGTCA CACAGAGTAC


         1908        1918       1928        1938        1948
GCAAACACAC CGATGCACAC GCACATAGAG ATATGCACAC ACAGATGCAC
```

```
      1958        1968        1978        1988        1998
ACACAGATAT ACACATGGAG TGCACGCACA TGCCAATGCA CGCACACATC


      2008        2018        2028        2038        2048
AGTGCACACG GATGCACAGA GATATGCACA CACCGATGTG CGCACACACA


      2058        2068        2078        2088        2098
GATATGCACA CACATGGATG AGCACACACA CACCAAGTGC GCACACACAC


      2108        2118        2128        2138        2148
CGATGTACAC ACAGATGCAC ACACAGATGC ACACACACCG ATGCTGACTC


      2158        2168        2178        2188        2198
CATGTGTGCT GTCCTCTGAA GGCGGTTGTT TAGCTCTCAC TTTTCTGGTT


      2208        2218        2228        2238        2248
CTTATCCATT ATCATCTTCA CTTCAGACAA TTCAGAAGCA TCACCATGCA


      2258        2268        2278        2288        2298
TGGTGGCGAA TGCCCCCAAA CTCTCCCCCA AATGTATTTC TCCCTTCGCT


      2308        2318        2328        2338        2348
GGGTGCCGGG CTGCACAGAC TATTCCCCAC CTGCTTCCCA GCTTCACAAT


      2358        2368        2378        2388        2398
AAACGGCTGC GTCTCCTCGC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA


      2408        2418        2428        2438
AAAAAAAAAA AAGGAATTCG AGCTCGGTAC CCGGGGATCC
```
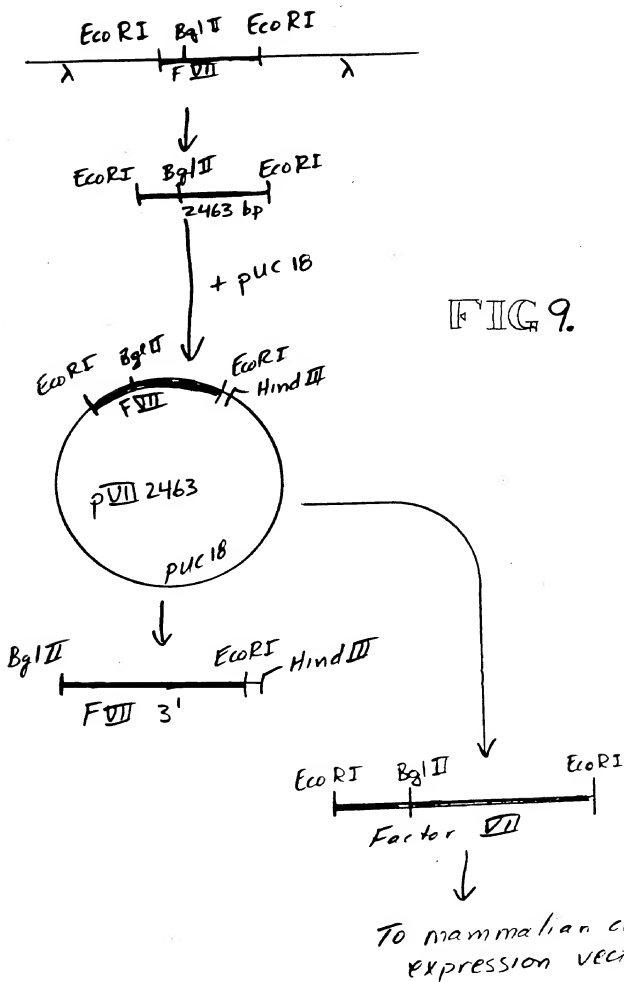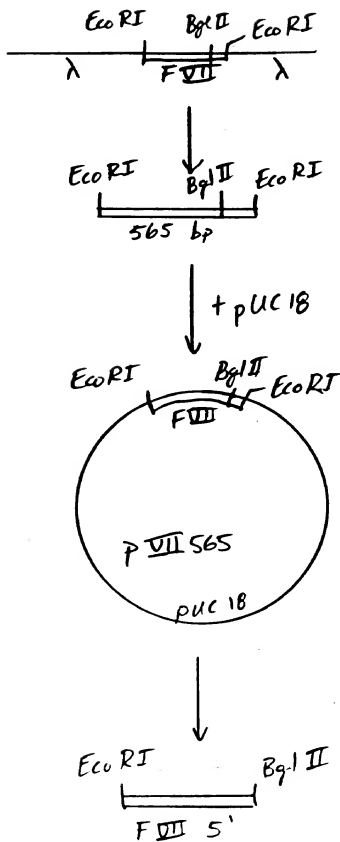
FIG 9.
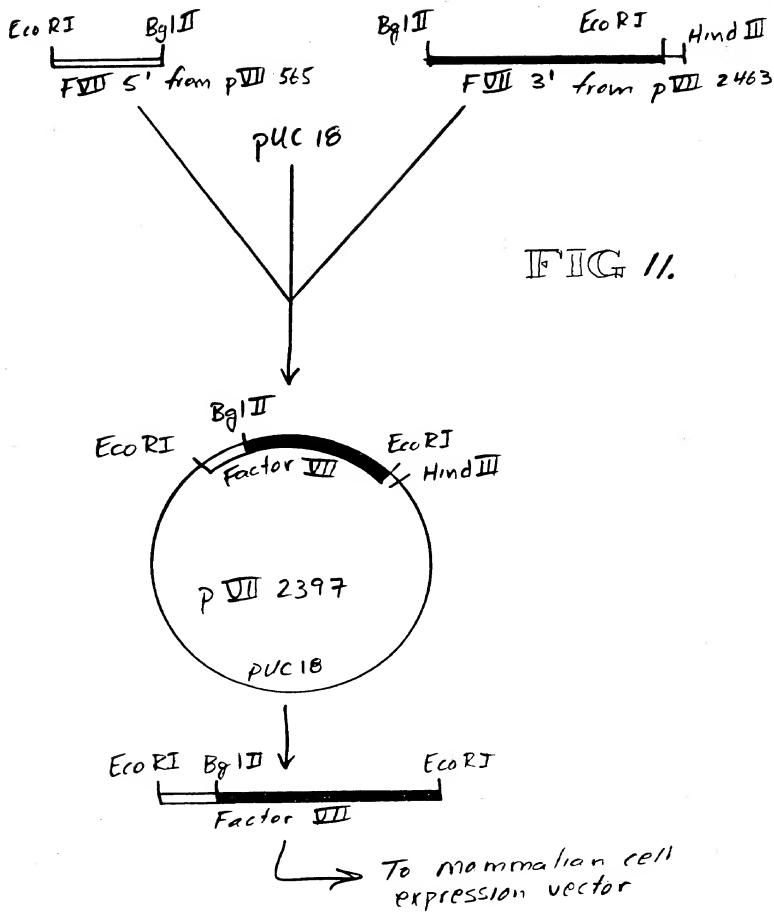
To mammalian cell expression vector

FIG 10.

FIG 11.
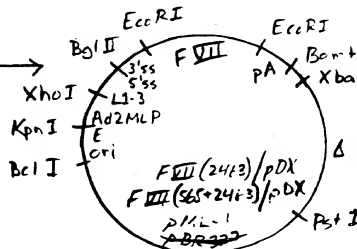
FIGURE 12
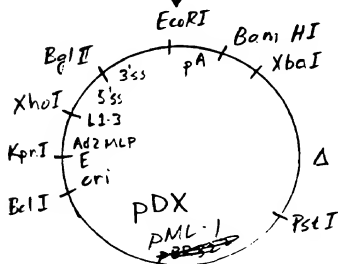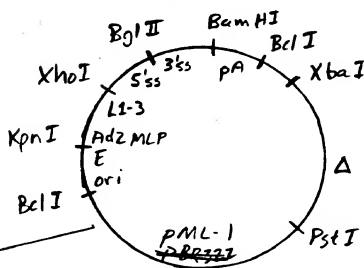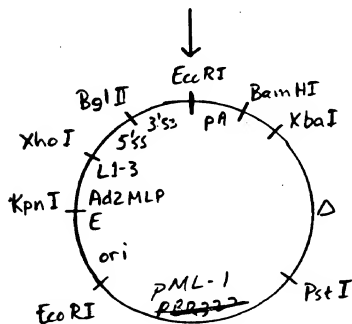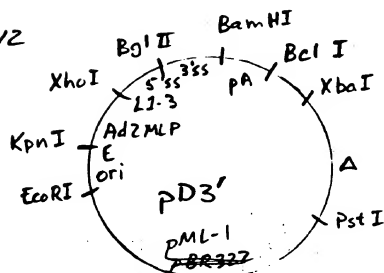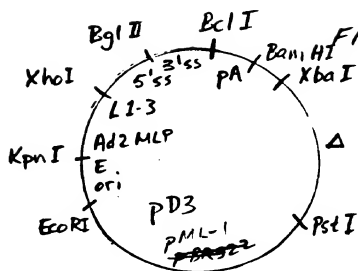
```
          714                    729                744
CGG ACG TTC TCT GAG AGG ACG CTG GCC TTC GTG CGC TTC TCA TTG GTC AGC GGC
Arg Thr Phe Ser Glu Arg Thr Leu Ala Phe Val Arg Phe Ser Leu Val Ser Gly

759                774      Nar I     789                  804
TGG GGC CAG CTG CTG GAC CGT GGC GCC ACG GCC CTG GAG CTC ATG GTC CTC AAC
Trp Gly Gln Leu Leu Asp Arg Gly Ala Thr Ala Leu Glu Leu Met Val Leu Asn

        819                    834      Pst Ib     849                864
GTG CCC CGG CTG ATG ACC CAG GAC TGC CTG CAG CAG TCA CGG AAG GTG GGA GAC
Val Pro Arg Leu Met Thr Gln Asp Cys Leu Gln Gln Ser Arg Lys Val Gly Asp

                879                    894                    909
TCC CCA AAT ATC ACG GAG TAC ATG TTC TGT GCC GGC TAC TCG GAT GGC AGC AAG
Ser Pro Asn Ile Thr Glu Tyr Met Phe Cys Ala Gly Tyr Ser Asp Gly Ser Lys

    924                    939                954                    969
GAC TCC TGC AAG GGG GAC AGT GGA GGC CCA CAT GCC ACC CAC TAC CGG GGC ACG
Asp Ser Cys Lys Gly Asp Ser Gly Gly Pro His Ala Thr His Tyr Arg Gly Thr

                984                    999                1014
TGG TAC CTG ACG GGC ATC GTC AGC TGG GGC CAG GGC TGC GCA ACC GTG GGC CAC
Trp Tyr Leu Thr Gly Ile Val Ser Trp Gly Gln Gly Cys Ala Thr Val Gly His

1029                    1044                1059 TaqI         1074
TTT GGG GTG TAC ACC AGG GTC TCC CAG TAC ATC GAG TGG CTG CAA AAG CTC ATG
Phe Gly Val Tyr Thr Arg Val Ser Gln Tyr Ile Glu Trp Leu Gln Lys Leu Met

        1089                    1104                1119                    1138
CGC TCA GAG CCA CGC CCA GGA GTC CTC CTG CGA GCC CCA TTT CCC TAG CCCAGCAGCC
Arg Ser Glu Pro Arg Pro Gly Val Leu Leu Arg Ala Pro Phe Pro
                                                                          PstIc
    1148         1158         1168         1178         1188         1198    1208
CTGGCCTGTG GAGAGAAAGC CAAGGCTGCG TCGAACTGTC CTGGCACCAA ATCCCATATA TTCTTCTGCA

        1218         1228         1238         1248         1258         1268    1278
GTTAATGGGG TAGAGGAGGG CATGGGAGGG AGGGAGAGGT GGGGAGGGAG ACAGAGACAG AAACAGAGAG

        1288         1298         1308         1318         1328         1338    1348
AGACAGAGAC AGAGAGAGAC TGAGGGAGAG ACTCTGAGGA CCATGGACAG AGACTCAAAG AGACTCCAAG

        1358         1368         1378         1388         1398         1408    1418
ATTCAAAGAG ACTAATAGAG ACACAGAGAT GGAATAGAAA AGATGAGAGG CAGAGGCAGA CAGGCGCTGG

        1428         1438         1448         1458         1468         1478    1488
ACAGAGGGGC AGGGGAGTGC CAAGGTTGTC CTGGAGGCAG ACAGCCCAGC TGAGCCTCCT TACCTCCCTT
```

```
        1498        1508        1518        1528        1538        1548        1558
CAGCCAAGCC CCACCTGCAC GTGATCTGCT GGCCCTCAGG CTGCTGCTCT GCCTTCATTG CTGGAGACAG

        1568        1578        1588        1598        1608        1618        1628
TAGAGGCATG ACACACATGG ATGCACACAC ACACACGCCA TGCACACACA CAGAGATATG CACACACACG

        1638        1648        1658        1668        1678        1688        1698
GATGCACACA CAGATGGTCA CACAGAGTAC GCAAACACAC CGATGCACAC GCACATAGAG ATATGCACAC

        1708        1718        1728        1738        1748        1758        1768
ACAGATGCAC ACACAGATAT ACACATGGAG TGCACGCACA TGCCAATGCA CGCACACATC AGTGCACACG

        1778        1788        1798        1808        1818        1828        1838
GATGCACAGA GATATGCACA CACCGATGTG CGCACACACA GATATGCACA CACATGGATG AGCACACACA

        1848        1858        1868        1878        1888        1898        1908
CACCAAGTGC GCACACACAC CGATGTACAC ACAGATGCAC ACACAGATGC ACACACACCG ATGCTGACTC

        1918        1928        1938        1948        1958        1968        1978
CATGTGTGCT GTCCTCTGAA GGCGGTTGTT TAGCTCTCAC TTTTCTGGTT CTTATCCATT ATCATCTTCA

        1988        1998        2008        2018        2028        2038        2048
CTTCAGACAA TTCAGAAGCA TCACCATGCA TGGTGGCGAA TGCCCCCAAA CTCTCCCCCA AATGTATTTC

        2058        2068        2078        2088        2098        2108        2118
TCCCTTCGCT GGGTGCCGGG CTGCACAGAC TATTCCCCAC CTGCTTCCCA GCTTCACAAT AAACGGCTGC

        2128        2138        2148        2158        2168        EcoRIb
GTCTCCTCGC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAGGAATTC
```